Subband Speech Coding and Matched Convolutional Channel Coding for Mobile Radio Channels

Richard V. Cox, *Fellow, IEEE*, Joachim Hagenauer, *Senior Member, IEEE*, Nambirajan Seshadri, *Member, IEEE*, and Carl-Erik W. Sundberg, *Fellow, IEEE*

***Abstract*—Due to increased radio spectral congestion, the trend in future cellular mobile radio systems is toward digital transmission. The recent advances in spectrally efficient modulation techniques and high quality low bit rate speech coding have further aided this move. However, mobile radio channels are subject to signal fading and interference which causes significant transmission errors. The design of speech and channel coding for this application is therefore challenging. In this paper, the effects of digital transmission errors on a family of variable-rate embedded subband speech coders (SBC) have been analyzed in detail. It is shown that there is a difference in error sensitivity of four orders of magnitude between the most and the least sensitive bits of the speech coder. As a result, a family of rate-compatible punctured convolutional (RCPC) codes with flexible unequal error protection capabilities have been matched to the speech coder. These codes are optimally decoded with the Viterbi algorithm. On a Rayleigh fading channel with differential four phase shift keyed modulation, more than 5 dB gain in channel signal-to-noise ratio can be obtained by using 4 levels of unequal error protection over conventional designs that utilize only 2 levels. This gain is achieved over a large range of channel signal-to-noise ratios, at no extra bandwidth requirement and only a small complexity increase. Among the results, analysis and informal listening tests show that with a 4-level unequal error protection scheme, transmission of 12 kb/s speech is possible with very little degradation in quality over a 16 kb/s channel with an average bit error rate of $2 \cdot 10^{-2}$ at a vehicle speed of 60 mph and with interleaving over two 16 ms speech frames. The SBC speech encoder/decoder and the RCPC channel coder/decoder have been implemented on a single AT&T DSP-32 floating point signal processor. The overall end-to-end delay is about 88 ms.**

I. Introduction

Low and medium bit rate coders are currently being considered for deployment in future mobile radio systems [1], [2]. The main motivations for these are the efficient use of the radio spectrum and the ability to offer ISDN access facilities [3]. Advancements in high quality low bit rate speech coding algorithms with the associated ability to implement them on high speed digital signal processors [4], and the development of spectrally efficient digital transmission methods have further aided this move.

Manuscript received December 8, 1988; revised July 18, 1990. Parts of this paper were presented at ICASSP '88, New York, NY, and at the 38th Vehicular Technology Conference, Philadelphia, PA, June 1988.

R. V. Cox, N. Seshadri, and C.-E. W. Sundberg are with AT&T Bell Laboratories, Murray Hill, NJ 07974.

J. Hagenauer is with the German Aerospace Research Establishment (DLR), D-8031 Oberpfaffenhofen, Germany.

IEEE Log Number 9100764.

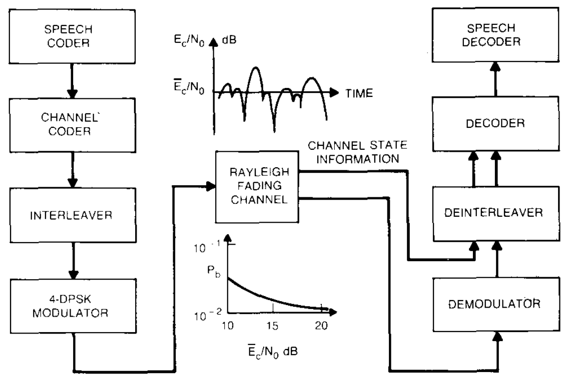


Fig. 1. Block diagram of the communication system.

Benefits of a digital technology include reliable encryption for secure voice transmission.

A block diagram of the system considered in this paper is shown in Fig. 1. The modulator considered is quaternary phase shift keying (4-PSK) with differential encoding and differentially coherent detection (4-DPSK). This figure also shows the variability of the channel signal-to-noise ratio due to Rayleigh fading. We have also shown a typical worst case curve for the average bit error probability versus the average signal-to-noise ratio for this Rayleigh fading channel [5] with 4-DPSK modulation. It can be seen that a major issue facing digitized speech transmission over a mobile radio channel is the impact of channel fading which results in error rates between 1% and 5%. Left untreated, these high error rates result in a speech communication system that is of unacceptable quality. Hence, some form of highly efficient error control coding is necessary to mitigate the effect of channel errors.

The channel variation in a mobile environment depends on several factors, such as the position of the mobile in a coverage area, the speed of the mobile, etc. Depending on the local channel characteristics, one would like to change the allocation of the bits for speech and channel coding. On a good channel, fewer bits should be spent on channel coding than on a noisy channel so that the overall noise due to both quantization and the channel distortion is minimized. Hence there is a requirement for the speech coder to operate at several different bit rates.

1053-587X/91/0800-1717$01.00

In this work, we consider the design of a flexible and adaptive speech coding system for the mobile radio channel. The speech coder is a dynamic bit allocation subband coder (D-SBC) which is capable of providing communication quality speech at a lower end of 10 kb/s and high quality speech at a high rate of 16 kb/s [6]. A novel set of embedded nonlinear quantizers enables the same coder and decoder to operate at all these bit rates.

The design of an error correction scheme usually consists of selecting a fixed channel code with a certain rate, complexity, and a correction capability that is uniform for all the data to be transmitted. Further, the fixed code is constructed for the worst case of average channel or source conditions. However, to make the best available use of the limited channel bandwidth, it is necessary to match the error protection provided by the channel code to the error sensitivity of the specific speech coder. For example, in our coder, we have noticed that error rates of the order of 1% can be tolerated easily for the least significant bits of the bit stream. Much more stringent requirements are necessary for the more significant bits. Similar observations have been made about other source coding systems [7]–[10]. Clearly, it would be a waste of bits to provide uniform error protection for the entire bit stream based only on the error sensitivity of the more significant bits. On the other hand, if the basis for error protection is the error sensitivity of the less significant bits, then the more sensitive ones will be exposed to an unacceptably high error rate. High quality speech coders are typically block processors and are made adaptive in order to exploit the perceptual properties of the human ear. The task of determining the relative importance of the bits in the bit stream and then constructing an error correcting code to match these measurements is quite challenging.

A block channel coding scheme would seem natural for error protection since these sophisticated speech coding schemes also generate blocks of digitized speech. Unequal error protection block codes have been studied in literature following the work of Masnick and Wolf [11]. However, these algebraic block coding schemes cannot easily accommodate broadly varying unequal error protection needs within one code word. Different portions of the encoded speech frame can of course be encoded and decoded with different block codes but this leads to either very short and inefficient codes or too long a delay. Furthermore, block codes are easily decodable only if the demodulator output is a hard decision rather than a soft decision which leads to error performance degradation. To distinguish between hard and soft decisions, we observe that although the data to be transmitted is binary, the demodulated signal takes on an analog value. This is due to signal fading and additive noise. When binary transmission and hard decision demodulation is employed, the output of the demodulator is quantized to either a logical zero or one. By contrast, the soft demodulator consists of more than two values, and the occupancy of a value gives us the reliability of the demodulated output being a logical zero or one. One of the simplest soft demodulators is a 3-zone demodulator, in which two of the values are logical zero and one, and the third is an "erasure," which is the output when it is considered to be sufficiently unreliable to declare the demodulated value as logical one or zero. The best soft demodulator would deliver the received value in a floating point format.

On a fading channel, the potential reduction in transmitted signal power is considerable when the decoder can use soft decisions rather than hard decisions [5]. Further improvement can be obtained if the time variation of the channel amplitude (channel state information, CSI) can be used in decoding. Hence, in this work, we propose the use of rate-compatible punctured convolutional (RCPC) codes [12], since they meet the requirements of handling soft decisions for decoding, as well as being capable of providing a flexible and wide range of unequal error protection. The RCPC codes are easily decodable by a maximum likelihood sequence estimator which can be efficiently implemented by the Viterbi algorithm [13]. The time variation of the channel amplitude can also be incorporated easily into the Viterbi algorithm to enhance the reliability of decoding. This cannot be readily accomplished with algebraic decoding techniques for block codes. Another significant advantage of the RCPC codes is that only one decoder is needed even if the level of error protection, and hence the code rate, changes several times within one speech frame. Although convolutional codes of this type are ideally suited to handle uncorrelated errors, errors which occur in bursts can be pseudorandomized through some form of interleaving (at the expense of added delay).

Combined source and channel coding schemes of the above type are evaluated in this paper through a mixture of analysis and simulations. In so doing we will make use of the developments and results on subband speech coding and convolutional coding from [6] and [12] and the error sensitivity analysis techniques in [7]–[10]. We have generalized and refined these techniques so that the sensitivity to transmission errors in individual bits in a block of speech can be calculated. Segmental signal-to-noise ratio is used to measure the objective quality of speech in the presence of both quantization and transmission noise. For tutorial texts on speech and channel coding we refer the reader to [14] and [13], respectively. The performance of RCPC codes in the narrow-band mobile radio channel that we are considering is presented in detail in the companion paper [5]. The details there include a complete description of the RCPC codes, the mobile radio channel model, and the theoretical and simulated bit error rate performance. We shall make use of these bit error rate results in evaluating our combined speech and channel coding systems.

For systems with high performance speech coders and advanced channel coding it is very time consuming to perform computer simulations for a wide variety of signal design parameters and channel conditions. For the purpose of simplifying the task of performance evaluation of such systems, we present a general method of calculating the segmental speech signal-to-noise ratio. The formula

takes as inputs a set of error sensitivity parameters, which are calculated once for a particular speech coder. Another set of input parameters are the individual bit error probability values. These can be calculated for simple transmission systems, but in practice they are simulated for a particular channel, channel coder, and modulation system. With our approach we have in effect split the overall system simulations in component source and channel simulations with a huge saving in computer time as a result. This cannot be done without approximations. Thus, our method should be used for initial screening, comparison and rough optimization of, e.g., coding and modulation designs.

In Section II, we present a general method to analyze the effect of transmission errors on a speech coding system. Based on this analysis, the error sensitivity of the bitstream is obtained. In Section III, the optimum error protection levels are synthesized from the bit error sensitivity results so that a required segmental SNR is obtained at the receiver. In Section IV, we describe the subband coder used in this work, and the bit error sensitivity results are obtained for this specific coder in Section V. Section VI briefly describes the rate compatible punctured codes, the channel model, the optimal decoding algorithm, and summarizes the design methodology of the combined speech and channel coding system. Section VII describes the DSP implementation and Section VIII has performance evaluations of various systems including delay considerations. Section IX concludes our work with discussions.

II. Effect of Transmission Errors on the Speech Coder Using Segmental SNR

The goal of this section is to derive an objective measure of the error sensitivity of the digitized speech. This is done by injecting errors in each bit position of the digitized speech bit stream, and measuring the distortion due to that error. This is done in every frame, and the average distortion is defined on a segmental basis. The analysis generalizes those of [7]–[9] on pulse code modulation (PCM), and those of [7], [10] on differential PCM (DPCM). Our methodology is general enough to encompass a wide range of coders, and the analysis can be modified to suit different criteria.

A. Error Sensitivity Analysis

Let $x(t)$ be the speech waveform which is divided into J nonoverlapping segments of duration T_B seconds and then encoded into N bits. Let $q(t)$ be the reconstruction of the speech waveform at the decoder when there are no transmission errors. The encoding procedure is blockwise memoryless and the effect of a transmission error is mostly confined to that frame alone. We ignore the processing and transmission delays in these mathematical representations. In a real system, a delayed version of $x(t)$ should be compared to $q(t)$ so that the same part of the signals $x(t)$ and $q(t)$ are compared. The N-bit packet is then encoded by the channel coder into $N + N_R$ bits, modulated, and transmitted over the channel. At the receiver, the signal is demodulated and decoded into one of the 2^N possible binary information sequences. The modulo 2 difference between the actual and the decoded sequence is the error sequence. Let e be an integer representation of an error pattern (there are 2^N error patterns) with the single bit errors numbered 1 through N followed by two bit errors, etc. Let $q_e(t)$ be the reconstructed version of $x(t)$ when the error pattern e occurs across the N-bit packet. The number of such error patterns is 2^N and $e = 0$ corresponds to correct transmission. The speech signal power in block j is

$$p^{(j)} = \frac{1}{T_B} \int_{(j-1)T_B}^{jT_B} [x(t)]^2 \, dt \tag{1}$$

and the expected value of the noise power is

$$n^{(j)} = \frac{1}{T_B} E\left\{ \int_{(j-1)T_B}^{jT_B} [x(t) - q_e(t)]^2 \, dt \right\} \tag{2}$$

where the expectation is taken over all possible channel error patterns with the probability of the occurrence of the error pattern e being P_e. Then the signal-to-noise ratio for block j is $p^{(j)}/n^{(j)}$ and the long-term segmental signal-to-noise ratio in decibels can be obtained as

$$\text{SEGSNR} = \frac{1}{J} \sum_{j=1}^{J} 10 \log \left(p^{(j)}/n^{(j)}\right) \tag{3}$$

where the logarithm here and henceforth is in base 10. We will study (3) more closely to determine the parameters that need to be calculated from the speech samples in order to evaluate the performance of the speech coder in the presence of transmission errors. It follows from (2) that for a certain error pattern e, the noise contribution can be rewritten as

$$\int_{(j-1)T_B}^{jT_B} [x(t) - q_e(t)]^2 \, dt = \int_{(j-1)T_B}^{jT_B} [a^2(t) + b^2(t) + c(t)] \, dt \tag{4}$$

where

$$a(t) = x(t) - q(t)$$
$$b(t) = q(t) - q_e(t)$$

and

$$c(t) = 2[x(t) - q(t)][q(t) - q_e(t)].$$

Here, $a(t)$ represents the quantization noise, $b(t)$ represents the contribution solely due to channel errors, and $c(t)$ is a mutual error term. For the systems considered in [7]–[9], this mutual term is approximately zero and thus neglected. For quantizers with few bits per sample, we found in [10] that it should be included. The second and the third terms in the right-hand side of (4) are zero if there are no transmission errors. The expected value of the total noise power assuming independence of error pat-

terns across frames is then

$$n^{(j)} = \frac{1}{T_B} \int_{(j-1)T_B}^{jT_B} a^2(t)\, dt + \sum_{e=1}^{2^N-1} P_e \left\{ \frac{1}{T_B} \int_{(j-1)T_B}^{jT_B} \{b^2(t) + c(t)\}\, dt \right\} \quad (5)$$

where P_e is the probability that channel error pattern e occurs. The signal-to-noise ratio for block j can then be conveniently rewritten as

$$p^{(j)}/n^{(j)} = 1 \bigg/ \left(Q^{(j)} + \sum_{e=1}^{2^N-1} P_e \mathcal{Q}_e^{(j)} \right) \quad (6)$$

where

$$Q^{(j)} = \frac{\int_{(j-1)T_B}^{jT_B} [x(t) - q(t)]^2\, dt}{\int_{(j-1)T_B}^{jT_B} [x(t)]^2\, dt} \quad (7)$$

is the quantization noise normalized with the speech signal power, and

$$\mathcal{Q}_e^{(j)} = \frac{\int_{(j-1)T_B}^{jT_B} \{[b(t)]^2 + c(t)\}\, dt}{\int_{(j-1)T_B}^{jT_B} [x(t)]^2\, dt} \quad (8)$$

is defined as the $\mathcal{Q}$-factor corresponding to block j and error pattern e. This $\mathcal{Q}$-factor is a short-term objective measure of the effect of a particular error pattern [8].

B. Approximations

Equation (6) requires the calculations of 2^N parameters ($2^N - 1$ $\mathcal{Q}$-factors and quantization noise) for each segment of speech and since N is typically over 100 b there will be computational problems. However, simplifications can be made under the assumption that the error protection by the channel code is strong enough that the most likely error event is that of a single bit error affecting a given frame. Hence, with independent single bit errors, we have

$$\sum_{e=1}^{2^N-1} P_e \mathcal{Q}_e^{(j)} \approx \sum_{i=1}^{N} P_i \mathcal{Q}_i^{(j)} \quad (9)$$

where P_i is the bit error probability in position i and where $\mathcal{Q}_i$ is the $\mathcal{Q}$-factor associated with that bit error. It is assumed that the first N error patterns correspond to single bit errors. Appendix A provides a more detailed treatment which includes the effect of double error patterns. There, we will also justify the single error approximation even for high values of the bit error probability for independent bit errors. Thus for unequal error protection, the segmental SNR can be approximated as

$$\text{SEGSNR} = \frac{1}{J} \sum_{j=1}^{J} 10 \log \left[\frac{1}{Q^{(j)} + \sum_{i=1}^{N} P_i \mathcal{Q}_i^{(j)}} \right]. \quad (10)$$

From the above expression, we note that by calculating $(N + 1)$ parameters (N $\mathcal{Q}$-factors and quantization noise) for each speech block, the segmental SNR can be evaluated for any set of bit error probabilities P_i, $i = 1, \cdots, N$.

From (10), we obtain the long-term comparisons of the sensitivity to single errors in different bits by setting the quantization noise $Q^{(j)}$ to zero and by injecting an error event with $P_i = 1$ for bit i and $P_j = 0$ for $j \neq i$. For this bit i, we define the average normalized noise power caused by an error in position i as

$$\frac{1}{J} \sum_{j=1}^{J} 10 \log \mathcal{Q}_i^{(j)}. \quad (11)$$

This value is thus a single parameter objective measure of the average error sensitivity of a particular bit.

III. Synthesis of Error Protection Levels From Error Sensitivity Measurements

Our ultimate design goal is to provide the highest quality speech communication system that is possible for the given set of channel conditions (at a practical complexity level). If the quality that is already obtained at the receiver without channel coding is considered sufficient then there is no need for additional error protection. However, in a situation such as ours, the channel error rate is high enough that the quality of speech at the receiver is clearly unacceptable. Moreover, the channel bandwidth is also severely limited, and hence the problem would be to find the highest speech quality that is achievable given these constraints. This requires that the error protection provided be matched to the bit error sensitivity requirements so that those bits which are the most sensitive to channel errors get the most protection.

To find the necessary unequal error protection that is needed, we will use the intuitive rule that all the speech bits after error protection should contribute the same to the average overall noise caused by the transmission errors. It is not practical to use 100's of levels of error protection since the task of channel code design would be impractical. Many of these bits may have the same or similar error sensitivities. We will therefore cluster those bits with similar $\mathcal{Q}$-factors and use the same error protection levels for all these bits. To guide us in the design, we define the block $\mathcal{Q}$-factor for cluster k during segment j as

$$\overline{\mathcal{Q}}_k^{(j)} = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathcal{Q}_i^{(j)}. \quad (12)$$

We require each cluster, after error protection, to contribute the same noise on an average as any other cluster.

This equal impact of errors is achieved by letting the error protection profile $\{P_k\}$ be such that

$$\frac{1}{J}\sum_{j=1}^{J} 10 \log\left[\frac{1}{P_k \overline{\alpha}_k^{(j)}}\right] = D = \text{constant}, \qquad \forall k. \quad (13)$$

This is equivalent to

$$P_k\left[\prod_{j=1}^{J} \overline{\alpha}_k^{(j)}\right]^{1/J} = D_1 = 10^{-D/10}, \qquad \forall k. \quad (14)$$

The quantity D_1 can be determined by setting an explicit limit on the noise contribution due to transmission errors. For example, we can impose the rule that the optimum profile should satisfy (14), and that the segmental transmission noise power equal the segmental quantization noise power (a fair criterion for allocation of the available bits between the speech and channel coder). Then, we get

$$\frac{1}{J}\sum_{j=1}^{J} 10 \log\left[\frac{1}{\sum_{k=1}^{K} P_k N_k \overline{\alpha}_k^{(j)}}\right] = \frac{1}{J}\sum_{j=1}^{J} 10 \log\left[\frac{1}{Q^{(j)}}\right]. \quad (15)$$

Here P_k is the probability of a bit being in error in group k. This yields the following closed form expression for D_1 in decibels:

$$D_1 \text{ (dB)} = \frac{10}{J}\sum_{j=1}^{J}\left[\log Q^{(j)} - \log \sum_{k=1}^{K} \cdot \left(\frac{N_k \overline{\alpha}_k^{(j)}}{\left[\prod_{i=1}^{J} \overline{\alpha}_k^{(i)}\right]^{1/f}}\right)\right]. \quad (16)$$

However, one may not be able to use this criterion in reality, because the speech coder would be required to have a certain subjective quality under noiseless conditions, and this would necessitate allocating a fixed portion of the total bit rate for speech coding. In practice, the error probabilities P_k's of the channel coder can be simulated for the given channel coding rate. The values of $\overline{\alpha}_k$ can also be measured for the speech codes. Finally, we seek that profile P_k for which the average impact of errors in every cluster is low and are approximately equal to each other.

IV. Subband Speech Coder

Subband coding of speech is a relatively mature form of waveform coding of speech. The speech signal is first divided into a number of subbands, which are then individually encoded. The underlying principle for the coder is that the bit allocation can be weighted so that those subbands with the most important information get most of the bits. The initial subband coders used fixed bit allocations based on the average spectrum of speech. Typical of this generation of coders was the one by Crochiere *et al.* [15]. In 1982, Ramstad introduced the idea of dynamically changing the bit allocation [16] based on the energy of each subband. Here, the bands with the most energy get most of the bits. More recently, Honda and Itakura [17], and Soong *et al.* [18] proposed dynamically assigning bits in both time and frequency. Their work produces very high quality speech. The complexity of these algorithms is very high, however.

In this work, we have considered a coder based on the idea proposed by Ramstad. A block diagram of the transmitter portion of this coder is shown in Fig. 2. The speech is divided into six 500-Hz-wide subbands and into 16-ms frames using GQMF filterbanks [19], [20]. Each of the subbands produces 16 samples per frame. Thus there are a total of 96 samples to quantize and they constitute the main information. Five bits are used to quantize the energy of the samples in each subband, thus totaling 30 b of side information. In addition, 2 b are allocated to the side information for the purpose of synchronization and/or signaling. Thus, the number of side information bits per 16-ms frame is 32 and they use up 2 kb/s of the total bit rate. The quantizer reconstruction levels are proportional to the square root of the energies, which gives us an estimate of the standard deviation for each of the bands. This estimate is available at both the transmitter and the receiver and is the basis for the quantization of the subband signals. Quantization is essentially logarithmic over a 72-dB range.

Bit allocation is derived from the quantized energies using an iterative procedure. At each iteration, 16 b (1 b per subband sample) are allocated to one of the subbands. Each iteration consists of finding the subband with the largest rms value, halving this value, storing the result in an rms table, and allocating 16 b to that subband. There is one additional proviso—no frequency band can be allocated more than 4 or 5 b per sample, depending on the maximum bit rate of the coder. Each iteration represents 1 kb/s of information. A nonuniform embedded quantizer optimized for a Gaussian input is used to quantize the individual subband samples. The step sizes of this quantizer are adjusted according to the quantized rms value of the subband. The principle behind the construction of the nonuniform quantizer is given in [6] and we will not go into the details. Embedded coding along with the concept of a prioritized bit stream enables the same coder and the decoder to operate at different rates in increments of 1 kb/s.

The key feature of our coder is the prioritized bit stream. This feature permits the rate of the coder to be changed in steps of 1 kb/s simply by "snipping" off the appropriate number of 16 b packets appearing in the end of a frame. Because of the embedded nature of the quantizer, the samples can be reconstructed at the decoder. In the mobile radio context this permits the appropriate division of the total bit rate that is available between the speech and the channel coder.

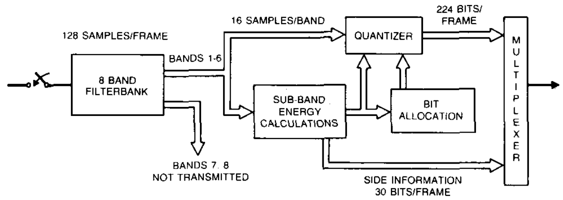


Fig. 2. Dynamic bit allocation subband coder.

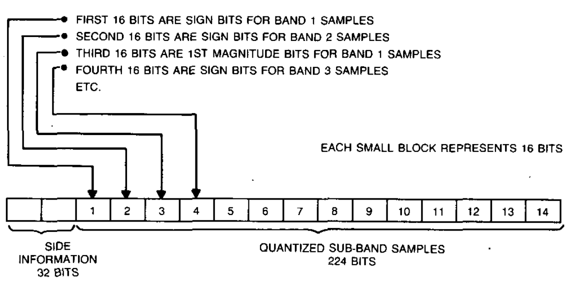


Fig. 3. Example of frame organization for 16 kb/s subband coder.

Perhaps the most important benefit of the prioritized bit stream for error protection comes from the fact that the average bit error sensitivity of the main information bits in a digitized speech packet decreases monotonically as we move from the beginning of the packet to the end thus lending themselves to unequal error protection. We will show that prioritization makes the relative importance of these bits almost invariant to input nonstationarity. This pseudostationarity of the error sensitivity is extremely crucial in designing the error protection since it places less stringent requirements on the channel code. A nonstationary bit error sensitivity profile would require us to change the channel coding adaptively, thus rendering the task of error protection more difficult. In particular, we have noted that the main information bit error sensitivity is invariant to changes in the input power level fluctuations.

Fig. 3 shows how the bit stream prioritization can be accomplished by considering a digitized speech packet for the 16 kb/s coder during a frame of 16 ms. There are 256 b to be allocated and they can be thought of as 16 b words. In this example, the order of bit allocation is subband 1, subband 2, subband 1, etc. The bit stream is also arranged in the same order. At the beginning of the stream are two 16 b words allocated for side information. The third 16 b word is allocated to subband 1 and they are the sign bits of the quantized samples for that band. The fourth 16 b word is allocated to subband 2 and they are the sign bits for subband 2 quantized samples. The fifth 16 b word is for the subband 1 samples and they correspond to the most significant magnitude bits for the samples of subband 1.

At the receiver, the side information is decoded first. Based on the side information, and with the knowledge of the rate of the coder, the bit allocation can be determined. For variable rate coding, the rate of the code can be transmitted as additional side information. From this knowledge, the remaining bits can be decoded to reconstruct the 96 subband samples. The synthesis of these subband samples with the synthesis filter bank yields the speech output.

V. Error Sensitivity Analysis for the Subband Speech Coder

A. *Effect of Single Errors on the Speech Coder*

Here, we use the theoretical developments in Section II to evaluate the average error sensitivity of the bits from the subband coder. From this error sensitivity profile, we determine the tolerable values of error probability on these bits. To determine the average relative importance of a bit output from the coder, an error is introduced into that bit alone in every frame and the resulting segmental SNR is determined. Comparison of the segmental SNR's of the bits with error determines their importance. In our work, about one thousand 16-ms frames of male and female speech were considered in the averaging process. However, for the numerical results in this paper, we use the following 2 sentences corresponding to a total of 256 frames. The sentences are as follows.

Male: "Her father failed many tests."

Female: "An icy wind raked the beach."

The speech was band limited between 100 and 3000 Hz, and sampled at 8 kHz to give the speech samples for our simulations. The speech samples are processed by the analysis and synthesis filters of the subband coder before evaluating the speech power in a segment. To obtain the quantized samples of speech, the speech samples are analyzed and encoded into a binary stream. At the decoder the bits are reconstructed into the appropriate subband samples, and then synthesized. The difference between the latter and the former speech samples gives the quantization noise alone. To evaluate the effect of channel errors, single errors are added modulo 2 to the binary bit stream representing the coder output. The binary sequence with error is then reconstructed into subband samples, and then synthesized to give the speech that is distorted by the subband coding process and by the channel errors.

Fig. 4 shows the result of these calculations for the 12 kb/s speech coder. The horizontal axis represents the information bits at the speech encoder output. The vertical axis represents the decoded speech segmental SNR. The first 30 b represent the side information. The results are arranged by bands and then the order of bits for each subband. The left-most bit (number 3) represents the most significant bit for subband 1. Bit number 32 represents the least significant bit for subband 6. It can be seen that the two most significant bits of each band have the most impact on performance. When this experiment was repeated for other rates, only the absolute value of the segmental SNR for each bit differed and the relative importance remained the same. Perhaps the most significant result is that there is about 23 dB degradation caused by an error

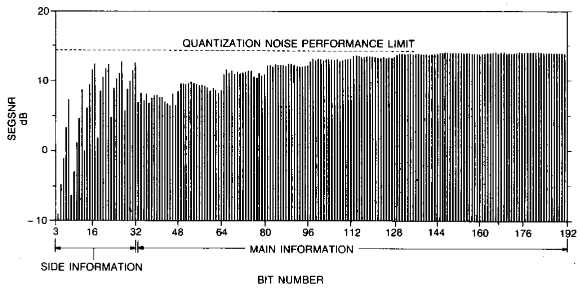


Fig. 4. Bit error sensitivity of the 12 kb/s coder. Single errors in bits 3 to 192.

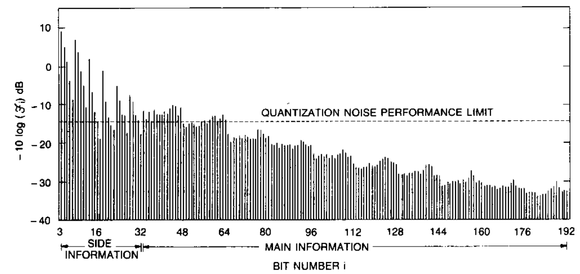


Fig. 5. Error sensitivity for the bits of the 12 kb/s coder calculated according to (11).

in the most significant bit of the side information when compared to the quantization noise limited performance which corresponds to about 14.2 dB. Another significant result is that there is more than 20 dB variation in error sensitivity even among the side information bits. This shows that some of the less significant side information bits are rather insensitive to channel errors. Fig. 4 also shows the result of the calculation performed for the main information bits. It can be seen that the first 16 b (numbers 33 to 48) are more sensitive to channel errors than the next 16 b, etc. The effect of errors on the least significant bits is almost negligible. Thus the subband coder main information output is inherently prioritized. However, it can be seen that some of the significant main information bits are more sensitive to channel errors than some bits of the side information. Thus, better prioritization can be obtained by rearranging some of the side information and main information bits. This would further facilitate unequal error protection.

In the subband coder we have used scalar quantizers with a sign bit code representation (first bit is a sign bit, next bit is the most significant magnitude bit, etc.). We also investigated the improvement in error sensitivity by using a better (more robust) bit representation (index assignment) for the quantizer output levels. For scalar quantizers we used the minimum distance code (MDC) representation [8] which is significantly more robust to transmission errors in PCM than the conventional sign bit code. For transmission without channel coding, especially for low level samples, there are significant improvements. However, with powerful channel coding for very noisy channels, the overall system performance difference using various bit representations (MDC or others) is minor.

B. Error Sensitivity

It can be seen from Fig. 4 that when an error occurs in the less significant bits, the noise due to channel errors is masked by the quantization noise. Hence, to obtain a clearer picture of how the errors in different bits contribute to the noise, we have calculated the error sensitivity without quantization noise as given by (11). We have observed that in some blocks j, and for certain bits i, $\mathcal{A}_i^{(j)}$ can be negative. This implies that inverting that bit actually decreases the total noise contribution for that block. This effect is mainly evident on the least significant bits of main information. Hence, we include only those segments contributing nonnegative single error $\mathcal{A}^{(j)}$ values. The long-term effect of this is plotted in Fig. 5 for the 12 kb/s speech coder on a decibel scale. Note that the variation of the noise level between the most and the least significant bits is more than 40 dB. Also note that for error free transmission, the normalized quantization noise is about -14.2 dB. The effect of transmission single errors on the least important bits are of little or no importance, since their individual sensitivity to errors is very low. However, if many of these less significant bits are in error, then their total noise contribution becomes significant. Thus omitting less significant bits or transmitting them with high error probability will soon have an impact on speech quality. The effect of double errors is discussed in Appendix A.

C. Segmental SNR with Ideal Error Protection

Fig. 6 illustrates the effect of the unprotected bits in a speech block on the overall segmental signal-to-noise ratio corresponding to the channel conditions of Fig. 1. The vertical axis shows the output segmental SNR. The horizontal axis is the average channel SNR per transmitted bit on a mobile radio channel [5]. The number of most significant 16 b subpackets that are perfectly protected is shown beside each curve. The segmental SNR is calculated using (10). The results show that for channel SNR's above 15 dB (bit error rate between 1% and 3%), at least the four most significant 16 b subpackets should be protected well. Otherwise the overall performance will be considerably degraded. These results serve as an upper bound to the speech quality that can be achieved using actual channel codes. Further, they also reveal how many bits can be left unprotected, and still have good quality speech.

VI. Channel Error Protection

A. Rate Compatible Punctured Codes

From earlier sections, we have seen that the bit stream from the speech encoder output exhibits widely varying error sensitivity within each speech frame. Hence, there

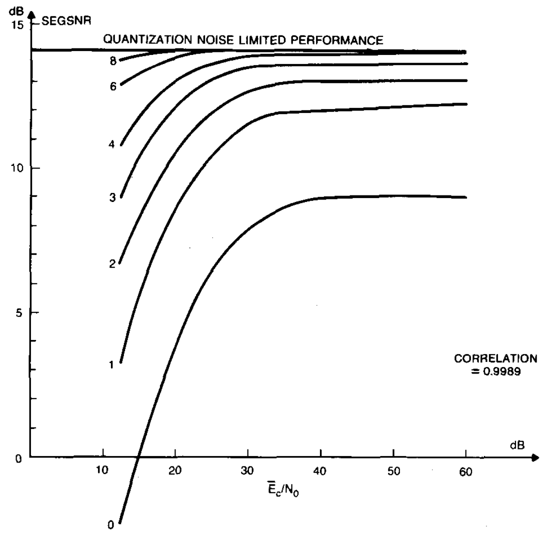


Fig. 6. Speech SEGSNR versus channel SNR for ideally protected 16 b subpackets.

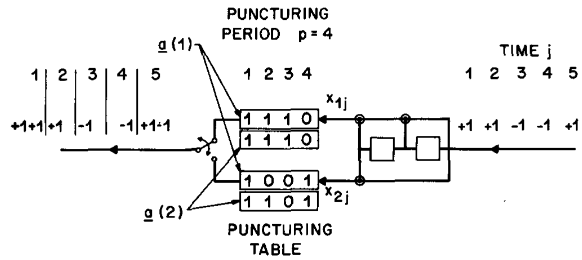


Fig. 7. Example of two RCPC codes with memory $M = 2$, puncturing period $p = 4$. Mother code rate $R = 1/2$. Punctured code rates are $R = 4/5$ and $R = 4/6$.

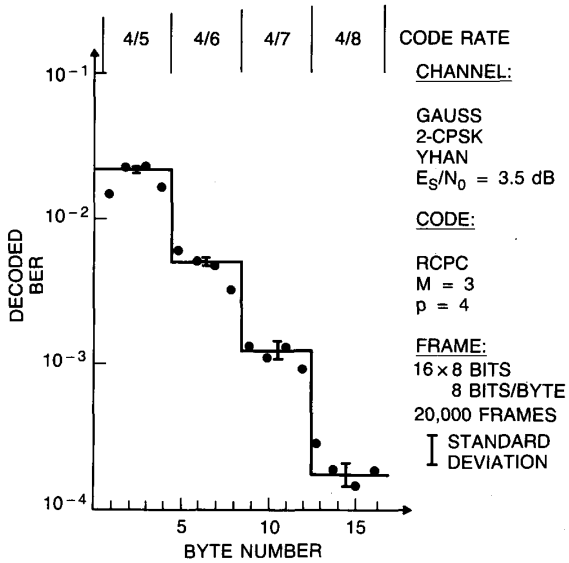


Fig. 8. Bit error rate for a RCPC code over the Gaussian channel with 4 different code rates. Memory $M = 3$, puncturing period $p = 4$. Gaussian channel and binary PSK with hard decisions.

is a need to build error correction codes that match this widely varying error sensitivity. Let us assume that there are K groups of bits, with the bits in group k requiring a final bit error rate of P_k. One could separately encode each of the K groups with K different encoders and decode with K different decoders. This has indeed been done by Suda and Miki [21] for error protection of backward adapted predictive speech coders. However, this configuration results in increased decoding complexity due to the use of different decoders. Further, short block codes are inherently inferior and very low channel code rates are needed to get the necessary error protection. This results in increased overhead. Here, we wish to use one single channel coder with a single maximum likelihood decoder which provides the error protection requirements within one speech frame with minimal amount of redundancy. This can be achieved when the concept of punctured convolutional codes [22] is modified by introducing a rate compatible restriction to the puncturing rule [12].

In order to briefly explain rate compatible punctured convolutional codes (RCPC), we start with the example of Fig. 7 where a rate $R = 1/2$ convolutional code with memory $M = 2$ is shown. The input to the convolutional encoder at any time j takes on values ± 1. The outputs of the encoder x_{1j} and x_{2j} also take the same values. These output symbols are punctured periodically with period $p = 4$ according to the puncturing rules $\boldsymbol{a}(1)$ or $\boldsymbol{a}(2)$ which are shown in the figure. A zero in the puncturing table means that the code symbol is not to be transmitted. Thus, if $\boldsymbol{a}(1)$ were used for puncturing, the outputs after puncturing would correspond to those in Fig. 7. At time $j = 1$, both the output bits are transmitted. At times $j = 2$, 3, and 4 only x_{12}, x_{13}, and x_{24} are transmitted. The puncturing rule is then repeated periodically. Thus, we have realized a rate $R = 4/5$ code by puncturing the lower rate $R = 1/2$ mother code. The puncturing table can be described by an $N \times p = 2 \times 4$ matrix:

$$\boldsymbol{a}(1) = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}. \tag{17}$$

To realize a rate $R = 2/3$ code which has a rate lower than $R = 4/5$ but higher than $R = 1/2$, one could use any puncturing rule that deletes the appropriate number of bits from the $R = 1/2$ mother code. However, if this code has to be compatible with the higher rate $R = 4/5$ code, it should also transmit the same bits as those transmitted by the $R = 4/5$ code. Thus, 1's in the $\boldsymbol{a}(1)$ puncturing matrix corresponding to $R = 4/5$ code cannot be changed to 0. In order to meet the rate compatibility requirement, we can use the puncturing table

$$\boldsymbol{a}(2) = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}. \tag{18}$$

In general, from a mother code of rate $R = 1/N$, we

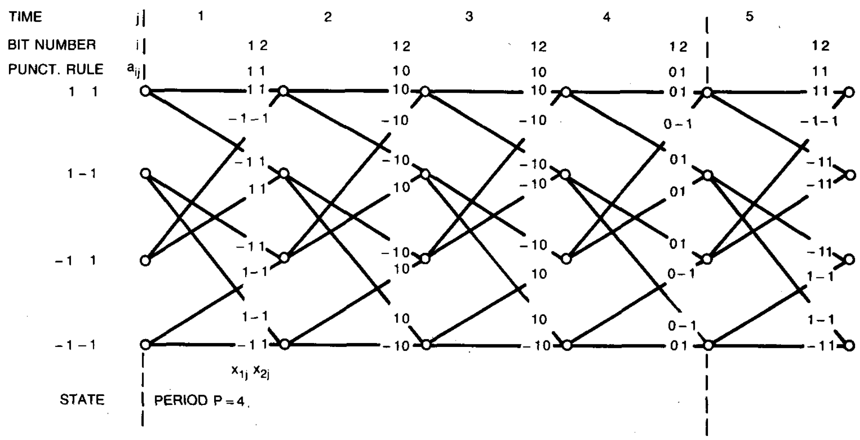


Fig. 9. Decoder trellis for the $R = 4/5$ RCPC code of Fig. 7.

can obtain a family of codes with rates

$$R = \frac{p}{p + l}, \quad l = 1, \cdots, (N - 1)p \tag{19}$$

where p is the puncturing period.

The reason for rate compatibility is explained in detail in [5], [12]. However, a brief explanation can be given with the aid of Fig. 8. Here, the simulated bit error probability performance of these codes are shown on a Gaussian channel with ideal coherent binary phase shift keying. In a transitional phase, as we go from a high rate code (example 4/5) to a low rate code (example 4/6), we have to guarantee that the error performance in the transition region does not degrade due to the influence of one code over the other. This is usually done in practice by terminating the code memory for each rate by using known tail bits. However, this results in a waste of redundancy. Alternatively, one could satisfy the rate compatibility condition to ensure that the transitional performance is at least as good as the performance of the high rate code, and typically much better. In the example, we note that the bit error probability of the high rate code near the transition region is better than its nominal performance (which is in the middle). This is due to the influence of the lower rate code which overall has a lower error probability. On the other hand, the performance of the lower rate code at the transition region suffers slightly due to the influence of the higher rate code.

B. The Viterbi Decoding Algorithm for Fading Channels

Optimal decoding of the transmitted signals in fading and noise can be efficiently performed by the Viterbi algorithm. Basically, the Viterbi algorithm (VA) for the memory $M = 2$ channel code in Fig. 7 operates on the trellis in Fig. 9. The rate of the code after puncturing is $R = 4/5$. The VA efficiently calculates the maximum likelihood sequence metric which is given by

$$\max_m \sum_{j=1}^{J} \lambda_j^m \tag{20}$$

where λ_j^m is the metric increment for path m in the trellis at time j. The maximum is evaluated over all 2^J possible paths, where J is the trellis depth. The metric increment is typically

$$\lambda_j^m = \sum_{i=1}^{N} a_{ij} x_{ij}^m y_{ij} \tag{21}$$

where a_{ij}'s are the elements of the puncturing matrix, x_{ij}^m is the trellis symbol (± 1) on branch j at bit number i for the path m. The y_{ij}'s are the received values either in the form of hard or soft decisions, and ranging from binary to floating point representations. During a fade, the received values are less reliable than during a nonfading situation. Thus if the fade information (channel state information (CSI), i.e., the estimated channel amplitude) is available at the receiver, it can be easily incorporated into the Viterbi algorithm. Let c_{ij} be the CSI for the received symbol y_{ij}. The metric increment is given by

$$\lambda_j^m = \sum_{i=1}^{N} a_{ij} c_{ij} x_{ij}^m y_{ij}. \tag{22}$$

The CSI is readily available at the receiver AGC. Incorporating the CSI improves the decoder performance [5].

C. Channel Model for Mobile Radio

The details of the channel simulation model are explained in detail in [5]. Only the essentials are presented here. Differentially encoded and detected 4-PSK (4-DPSK) is chosen as the modulation format. The channel imposes correlative Rayleigh fading on the transmitted 4-DPSK symbol (each symbol is 2 b). The correlation in the fades is generated by coloring a white noise se-

quence with a finite impulse response (FIR) filter whose cutoff frequency is controlled by the carrier frequency, the symbol transmission rate and the vehicle speed. Slower vehicle speeds result in longer fades. This fact, combined with limited interleaving results in higher residual error rates after Viterbi decoding than if the vehicle speed were higher. All of our following results are given for 60 mph vehicle speed because the fades are very long at slower vehicle speeds, resulting in excessively long times for simulations.

D. Summary of the Code Design Methodology

The flowchart of Fig. 10 summarizes our design effort for the combined speech and channel coding system. The speech coder is first analyzed for its sensitivity to channel errors. This requires the evaluation of the $\mathcal{Q}$ factors as described in Section II. We then fix the output SNR under noisy channel conditions to be at a certain value below the clear channel speech quality (for example, 3 dB below the clear channel quality). This, combined with the error sensitivity analysis gives us the bit error probability profile that is needed, and the procedure to calculate the profile is presented in Section III. For a fixed set of channel conditions (type of fading, modulation, and interleaving (delay considerations)) and decoding complexity (code memory, soft or hard decisions, availability of CSI, etc.), we seek the lowest possible average channel SNR at which the system can operate to give the required output speech quality. This step requires an extensive knowledge of the channel code performance for various code rates, and this can be done for simple channels by analysis and for more difficult channels such as ours by simulations. An additional constraint we have is the total bit rate. If the redundancy required to meet the error protection requirements exceed the allocated bits for this purpose, then we need to either relax the output SNR requirements or increase the decoding complexity. Increasing the overall delay also helps in the form of larger interleavers which result in a better randomization of channel burstiness. It should be emphasized that the optimum error protection scheme varies with the channel SNR when everything else is kept unchanged.

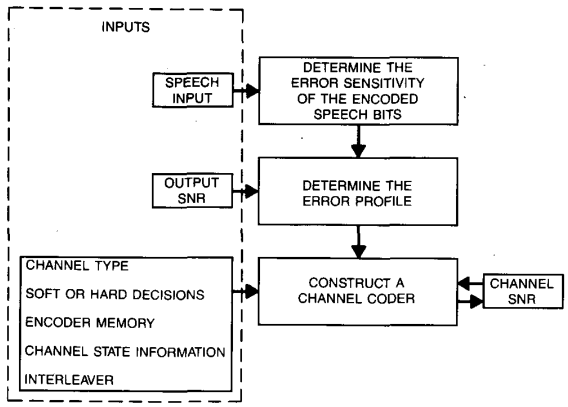


Fig. 10. Joint code design methodology.

E. Codes for 12 kb/s D-SBC Protection

We have designed two RCPC codes for the protection of the prioritized bitstream. The code itself has a memory M of 4. The puncturing period p is 8. Thus in the implementation of the Viterbi decoder, we have to calculate 4 incremental metrics, update the total metric into each of the 16 states, and update the path information.

The first code is designed for a "clean" channel with the uncoded channel bit error rate being 2%. The average channel SNR per bit $(\bar{E}_c/N_0)$ is about 17.5 dB. We have four error protection levels, where every 16 ms, the most important 16 bits are coded with the rate $R = 1/2$ mother code. The next 48 b are coded with the rate $R = 2/3$ code. The next 64 b are coded with the rate $R = 8/11$ code and the last 64 are left unprotected. This results in adding 16, 24, 24, and 0 b of redundancy to the 4 levels, respectively, giving an overall rate of 256 b every 16 ms or 16 kb/s. In order to facilitate interleaving, two frames of speech bits are combined at a time. Further, we also note that there are 2 b per frame that are not used in speech coding. These 4 b per two frames can be used for terminating the memory of the convolutional code by transmitting a known 4-b pattern. Although the convolutional code is suitable for transmitting a continuous stream of bits, by terminating the memory, we have realized a "framed convolutional code." In this example, the Viterbi decoder uses hard decisions on the received symbols. Soft values of the channel state information are used in the decoding process. The detailed simulation results presented in [5] are summarized here in Table I.

The second code is designed for even noisier channels with the channel bit error rate being in the order of 4% to 6% ($\bar{E}_c/N_0$ is about 12-9 dB). Here, the most important 16 b are protected by a rate $R = 1/2$ code, the next 32 b with rate $R = 4/7$ code, the next 48 b with rate $R = 2/3$ code, and the last 96 b are left unprotected. This results in a total bit rate of 16 kb/s. It also turns out that this design is ideally suited for implementing both the D-SBC coder (full duplex) and the channel encoder and decoder into one single AT&T DSP-32 processor. In this example, the Viterbi decoder uses only the soft decisions on the received symbols. The channel state information is not used. The performance of this code is presented in Table II. In general, it is advantageous to use soft decision decoding. In addition, if CSI is also available, it fetches additional benefit.

VII. Implementation

Both the D-SBC speech coder and the RCPC channel coder were implemented full-duplex on a single AT&T DSP32 signal processor. The DSP32 is a floating point signal processor which comes in both 40 and 100 pin packages. The 40 pin package cannot address off-chip memory. The 100 pin package DSP32 was used because the two programs combined exceeded both the 4-kilobytes RAM and 2-kilobytes ROM limitations of the 40 pin package. Also, the newer 25 MHz DSP32 had to be used to provide sufficient processing capability for both algo-

TABLE I

BIT-ERROR-RATE PERFORMANCE OF FIRST RCPC CODE DESIGN ON CORRELATED RAYLEIGH FADING CHANNEL WITH 4-DPSK MODULATION

Channel SNR	17 dB	15 dB	12 dB
Unprotected bits (64)	2.00×10^{-2}	2.80×10^{-2}	3.89×10^{-2}
Rate 8/11 code (64)	2.00×10^{-3}	4.5×10^{-3}	1.70×10^{-2}
Rate 2/3 code (48)	3.00×10^{-4}	1.00×10^{-3}	6.00×10^{-3}
Rate 1/2 code (16)	1.5×10^{-5}	6.00×10^{-5}	4.0×10^{-4}

Vehicle speed = 60 mph, carrier frequency = 900 MHz. Simulation over 5000, 16 ms frames. Two frame interleaving. Hard decision demodulation and full soft channel state information in Viterbi decoding. The number of bits protected by each code is indicated in parentheses.

TABLE II

BIT-ERROR-RATE PERFORMANCE OF SECOND RCPC CODE DESIGN ON CORRELATED RAYLEIGH FADING CHANNEL WITH 4-DPSK MODULATION

Channel SNR	17 dB	12 dB	9 dB
Unprotected bits (96)	2.00×10^{-2}	3.89×10^{-2}	6.35×10^{-2}
Rate 2/3 code (48)	3.00×10^{-5}	2.60×10^{-3}	2.03×10^{-2}
Rate 4/7 code (32)	2.00×10^{-5}	6.00×10^{-4}	6.50×10^{-3}
Rate 1/2 code (16)	0.0	7.00×10^{-5}	1.30×10^{-3}

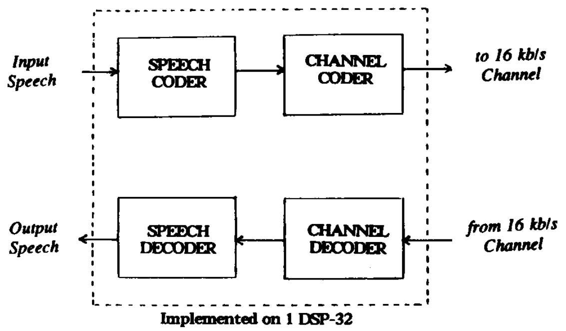
Vehicle speed = 60 mph, carrier frequency = 900 MHz. Simulation over 1000, 16 ms speech frames. Two frame interleaving. Soft decision demodulation and no channel state information in Viterbi decoding. The number of bits protected by each code is indicated in parentheses.

rithms, although either one alone could have fit on the older 16-MHz DSP-32. Approximately 53% of the real-time cycles were devoted to the speech coder and the remainder to the channel coder. The largest number of cycles were used in the Viterbi decoder. This algorithm required the simultaneous update of 16 states in the trellis as well as keeping track of the optimal path through each of the 16 states.

Fig. 11 is a block diagram of how the combined coders were implemented. At the transmitter the speech data from the codec A/D is first processed by the D-SBC encoder. The bitstream from this coder is then input to the RCPC encoder. The output bitstream from the RCPC encoder is then ready for the modulation system.

The DSP32 has two full-duplex ports, one which is serial and the other parallel. The serial port is occupied by the speech codec. The parallel port can be connected to a microprocessor which can manage the 16 kb/s datastream. Alternatively, dual-ported RAM could be used to transfer data between the channel and the coder.

At the decoder the 16 kb/s data stream is input to the RCPC decoder in a soft decision format. Bit values are represented on a scale of -1.0 to $+1.0$, with those values representing full confidence in the bits. Bits which have less confidence have a smaller magnitude. Bits which have been punctured are represented by a 0. The entire soft decision bit stream is then processed with the Viterbi decoder to determine the maximum likelihood path through the trellis. After processing all the RCPC protected bits in the frame, the state of the convolutional coder should be back to its original state, i.e., that formed by the four free bits we are using to "frame" the code. Thus, the path



DSP32	6.25 Mips
12 kb/s Speech Coder	3.66 Mips
RCPC Encode/Decode	2.34 Mips

Fig. 11. Block diagram of the D-SBC and RCPC full-duplex on a single AT&T DSP-32.

through the trellis leading to this state is chosen as the maximum likelihood path. Once the path has been chosen, the bit stream can be released and the D-SBC decoder can produce the output speech.

VIII. OVERALL PERFORMANCE

The overall performance of the system has been evaluated for different combinations of speech and channel coder rates. Some typical results are presented in Figs. 12 and 13 and 14 for the sub-band coder with various error protection strategies and channel conditions.

A. Results with Unequal Error Protection for 12 kb/s Speech Coder

Fig. 12 shows the segmental signal-to-noise ratio for 4 different methods of transmitting 12 kb/s speech over the 16 kb/s mobile channel. We have assumed that the threshold for acceptable quality is at about 2–3 dB degradation in SEGSNR compared to error free transmission. For reference, curve *A* shows the result for no error protection. Curve *B* has "perfect rate 1/3 error protection" on the side information of the subband coder. The remaining speech bits are transmitted unprotected resulting in a scheme with two levels of unequal error protection. This curve is clearly an upper bound on the performance with any real rate 1/3 code. Curve *D* gives the segmental signal-to-noise ratio for the case of a rate 3/4 convolutional code with 16 states applied uniformly to all speech bits in the frame. This corresponds to one level of equal error protection. Finally, curve *C* shows a system with 4 levels of unequal error protection using the channel code in Table II. The same interleaving depth and decoder complexity is used for systems *C* and *D*. As we pointed out before, the bits in a block from the speech encoder have significantly different sensitivities to transmission errors. It can therefore be expected that improvements are obtained by carefully matching the error protection level to the source bits. A comparison of curves *C* and *D* clearly illustrates this. Informal listening tests convey the same

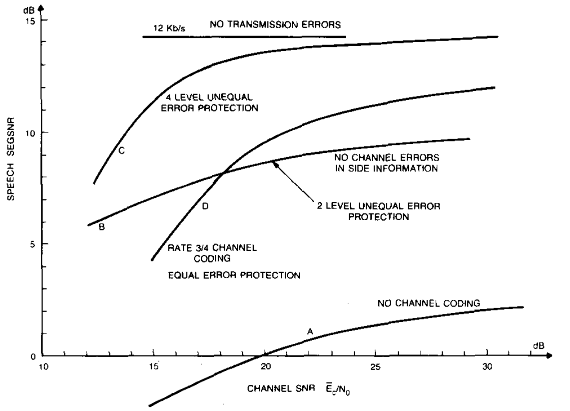


Fig. 12. Speech SEGSNR versus channel SNR for Rayleigh fading and 4 different transmission formats.

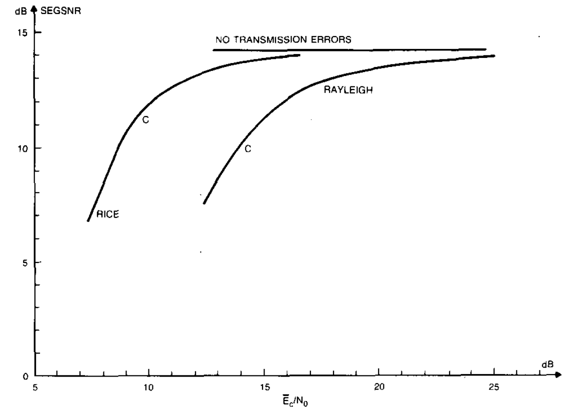


Fig. 13. Comparison of the performance of the 4-level protection system on Rayleigh and Rice channels; 12 kb/s speech and 2 frame interleaving.

message. For system B, it turns out that there are too many unprotected bits. As we will see later, it is better to use a less powerful code on a larger number of speech bits and leave fewer bits unprotected. Thus, with the same number of code states (and thus roughly the same system complexity) and identical transmission channel bandwidth, the matched 4-level system C clearly outperforms the equal error protection system in curve D.

Most of the discussions until now have centered around the problem of matching the source and channel code for the worst case Rayleigh fading channel. Implicitly, we have assumed that if the system can handle this case, it will also work for better transmission channels. This is illustrated in Fig. 13 which shows the performance of the same set of codes used in system C of Fig. 10 on a Rice channel with the Rice factor of 7 dB [5]. It is obvious that the performance on the Rice channel is much better. For an overall speech SEGSNR of 13 dB, under the Rice fading conditions, the average channel SNR can be as low as 9 dB.

B. Adaptive Assignment of Speech and Channel Coding Rates

In order to maximize the overall speech quality, it is necessary to adapt the speech and channel coding rates as a function of the channel condition. Under poor channel conditions, more bits should be spent on channel coding, which would reduce the quality of the speech coder (increased quantization noise), but would also reduce the distortion due to the channel (lower noise due to transmission errors), thus resulting in improved overall speech quality. Such an idea has been tried successfully with PCM (see [8], and further references therein) and ADPCM [23]. The embedded feature of the D-SBC and the flexibility of the RCPC channel coder makes this possible here too. Fig. 14 shows one such result for the current scenario. Under good channel conditions, 12 kb/s is used for speech coding and 4 kb/s for channel coding. Under worse conditions, the speech coder rate drops to 10 kb/s and the channel redundancy increases to 6 kb/s. We can thus see the considerable increase in the speech quality over a variety of channel conditions with this adaptive rate assignment. We have also shown in this figure, the results of performing interleaving over only one frame instead of the usual two. We can notice the considerable degradation in performance. However, if delay is of primary concern, then the figure shows that the speech coder rate should be dropped to 10 kb/s and that 6 kb/s of channel coding is a necessity.

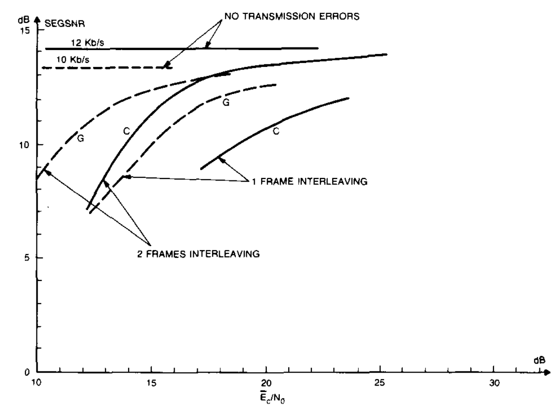


Fig. 14. Illustration of adaptive rate assignment with 12 and 10 kb/s speech coders.

C. Delay Budget

Fig. 15 shows the delay budget of the entire speech and channel coding system. The inherent delay (processor independent) of the speech coding system is 24 ms, of which 16 ms is due to the encoder/decoder buffers and 8 ms is due to the analysis/synthesis filter banks. The DSP-32 processing delay is about 16 ms per speech frame, which is the total time required to perform speech encoding, channel encoding, channel decoding (Viterbi decoding) and speech decoding. The time to transmit a protected speech frame of 256 b is 16 ms for a 16 kb/s modem. The receiver can start processing the data only after all the bits are received. This adds an additional delay of 16 ms. Thus, the total delay for a single frame interleaved speech and channel coding system is 56 ms.

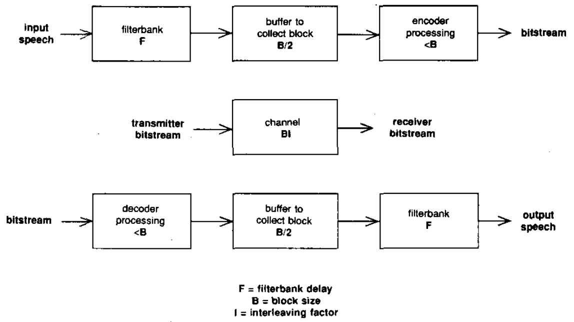


Fig. 15. Delay budget for the combined speech and channel coding system.

If interleaving is performed over more than a frame, additional delay is incurred. For interleaving over two frames, the first frame suffers an additional delay of 16 ms at the encoder as the second frame is being processed. Similarly, the second frame suffers a delay of 16 ms at the decoder as the first frame is being processed. In addition, the receiver needs to wait for 32 ms (transmission time of two frames) before it can start processing the received data. Thus, there is a delay of 48 ms solely due to two frame interleaving. The total delay is 88 ms. In general for interleaving over I frames, the interleaving delay alone is $(2I - 1)$ frames.

IX. Discussion and Conclusions

In this paper we have described a dynamic bit allocation subband coder and a rate compatible punctured convolutional channel coder. The subband coder produces good communications quality speech at bit rates as low as 12 kb/s. The coder also produces a prioritized bit stream which easily can be exploited for noisy channels. An analysis of the bit error sensitivities revealed that not all of the bits in the bit stream require error protection, and among those that do, unequal error protection is called for. The RCPC channel coder is extremely flexible and capable of producing unequal error protection while using the same convolutional encoder and decoder for all bits. When matched to the speech coder bit stream, a combined 16 kb/s coder was produced which was demonstrated to give robust performance over even fairly noisy channels. The technique which is used here, namely, to determine the error sensitivity of the subband coded bits and to match the channel coder according to the error sensitivity information, can be used with any other source coding scheme. Our main message in this paper is that there are large performance gains to be obtained in the system performance (with no increase in bandwidth and a small complexity increase) by carefully matching the source and channel coding.

Appendix A
Detailed Evaluation of the Effect of Transmission Errors

From (6), we can define the signal-to-noise ratio for block j as

$$\frac{p^{(j)}}{n^{(j)}} = \frac{1}{Q^j + \sum_{e=1}^{2^N-1} P_e \mathfrak{A}_e^{(j)}} \tag{A.1}$$

where $Q^{(j)}$ is the quantization noise which is defined in (7), and $\mathfrak{A}_e^{(j)}$ is the $\mathfrak{A}$ factor corresponding to an error pattern e for block j.

Equation (A.1) is impractical to evaluate for large N (N is more than 100 b for the subband coder). Instead, we consider approximations which analyze the effect of errors in terms of single and double error patterns. The probability of more than three errors (especially) after error protection is assumed to be negligible.

For independent, equiprobable bit errors with bit error probability P, the probability of a specific error pattern e with exactly w bit errors is

$$P_e = P^w(1 - P)^{N-w}. \tag{A.2}$$

This suggests that the contribution to the noise can be evaluated by grouping all the error patterns with the same weight, which has i. Indeed this has been done in [7]–[10]. Let us define the $\mathfrak{Z}$ factors through the relation

$$\sum_{e=1}^{2^N-1} P_e \mathfrak{A}_e^j = \sum_{w=1}^{N} P^w \mathfrak{Z}_w^{(j)}. \tag{A.3}$$

Let the first N $\mathfrak{A}$ factors ($e = 1, \cdots, N$) correspond to single bit error patterns. Then $\mathfrak{Z}_1$-factor for the jth segment is given by

$$\mathfrak{Z}_1^{(j)} = \sum_{i=1}^{N} \mathfrak{A}_i^{(j)} \tag{A.4}$$

and the $\mathfrak{Z}_2$ factor for the jth segment is obtained as

$$\mathfrak{Z}_2^{(j)} = \sum_{e'} \mathfrak{A}_{e'}^{(j)} - \mathrm{N}\mathfrak{Z}_1^{(j)} \tag{A.5}$$

where e' consists of all the $\binom{N}{2}$ double error patterns. Similarly, the higher order $\mathfrak{Z}$ factors can be obtained from the lower order $\mathfrak{Z}$ factors [9].

1. *Effect of Double Errors*

In the main text, the effect of single errors was studied in detail. However, in practice, even after channel error protection, more than one bit error can occur within a frame of speech, although with lower probability. To determine the effect of such errors, we consider the double bit error events with such error confined to one word. However, numerical results presented still assume that the bit errors are independent because of the inherent limitations in analyzing correlated errors. The segmental SNR based on the first two $\mathfrak{Z}$ factors is given by

$$\text{SEGSNR} \cong \frac{10}{J} \sum_{j=1}^{J} \log \left[\frac{1}{Q^{(j)} + P\mathfrak{Z}_1^{(j)} + P^2\mathfrak{Z}_2^{(j)}} \right]. \tag{A.6}$$

Tables III and IV summarize the $\mathfrak{Z}_1$ and $\mathfrak{Z}_2$ factors for a few segments of speech. The errors are confined to the

TABLE III
$\mathfrak{J}$-FACTORS FOR A SAMPLE OF SIX CONSECUTIVE HIGH ENERGY SEGMENTS OF SPEECH

Bits	$\mathfrak{J}_1$	$\mathfrak{J}_2$
Side Information Bits 3-16	9.66	−44.32
	5.92	−28.32
	5.43	−19.11
	4.82	−19.68
	7.01	−38.75
	4.39	−14.64
Main Information Bits 33-48	2.83	0.88
	2.59	0.71
	2.73	0.67
	2.63	0.82
	2.75	0.66
	2.96	1.01
Main Information Bits 160-176	1.02×10^{-2}	-3.17×10^{-5}
	8.90×10^{-3}	-2.52×10^{-5}
	1.06×10^{-2}	1.57×10^{-4}
	8.24×10^{-3}	-8.90×10^{-4}
	6.00×10^{-3}	3.30×10^{-5}
	1.22×10^{-3}	3.86×10^{-6}

TABLE IV
$\mathfrak{J}$-FACTORS FOR A SAMPLE OF SIX CONSECUTIVE LOW ENERGY SEGMENTS OF SPEECH

Bits	$\mathfrak{J}_1$	$\mathfrak{J}_2$
Side Information Bits 3-16	1.37×10^{4}	2.54×10^{5}
	8.86×10^{3}	1.62×10^{5}
	1.00×10^{4}	1.83×10^{5}
	1.61×10^{4}	2.97×10^{5}
	1.20×10^{4}	2.26×10^{5}
	9.80×10^{3}	1.61×10^{5}
Main Information Bits 33-48	0.65	6.57×10^{-2}
	0.53	2.92×10^{-2}
	7.47×10^{2}	1.64×10^{-2}
	0.44	-6.73×10^{-2}
	0.56	-5.68×10^{-2}
	0.56	0.20
Main Information Bits 160-176	3.38×10^{-3}	1.88×10^{-2}
	0.23	-8.25×10^{-2}
	0.25	-5.18×10^{-2}
	0.40	−0.24
	0.28	-3.85×10^{-2}
	0.32	0.15

side information bits 3-16, the main information (bits 33-48) and (bits 170-192). The speech coder is the 12 kb/s coder. We first consider the high energy segments for which the results are in Table III. These segments have their average energy considerably above the average energy of all the segments. Perhaps the most striking result is that the $\mathfrak{J}_2$ factors are in fact negative! This result means that the segmental SNR results using only single bit errors is pessimistic when the effect of double errors is also taken into account ($P^2\mathfrak{J}_2^{(j)}$ is negative). We next consider the 16 most significant main information bits (bits 33-48). For the same segments considered above, the $\mathfrak{J}_2$ factors are smaller than the $\mathfrak{J}_1$ factors resulting in negligible contribution to the noise power. The same effect is felt for the least significant bits. The effect of double errors is further mitigated by the fact that double errors are P times less probable than single errors (assuming independent errors), resulting in an average negligible noise power.

We next consider low energy segments in Table IV. These segments have their energy considerably below the average energy value of all the segments. The results are quite different as compared to high energy segments. Both $\mathfrak{J}_1$ and $\mathfrak{J}_2$ factors are considerably higher when there are errors in the side information. In fact $\mathfrak{J}_2$ factors are considerably higher than $\mathfrak{J}_1$ factors. This should be contrasted with the result for high energy segments where the $\mathfrak{J}_2$ factors are negative. The implication of this result is that double errors to the side information will be more noticeable during silent regions. Fortunately, the average noise power due to double errors is still small because of the less frequent occurrence (P^2) of these errors. The result also shows that the effect of errors in main information is negligible.

The above numerical example, together with the inherent interleaving in the speech coder/decoder seems to indicate that the single error analysis is a reasonable first-order approximation. In principle, we can easily evaluate the effect of error patterns with multiple errors produced by a channel decoder. This is done by injecting the most likely decoded error patterns, and evaluating the corresponding $\mathfrak{A}$ factors. However, this has to be repeated for every single channel encoder and decoder combination. The approximate single bit error method is independent of the channel code specifics.

REFERENCES

[1] *Proc. Second Nordic Seminar Digital Land Mobile Radio Commun.* (Stockholm, Sweden), Oct. 1986.

[2] J. H. Chen, G. G. Davidson, A. Gersho, and K. Zeger, "Speech coding for the mobile satellite experiment," in *Proc. IEEE Int. Conf. Commun. '87*, June 1987, pp. 756-763.

[3] E. S. K. Chien, D. J. Goodman, and J. E. Russel, Sr., "Cellular access digital network (CADN): Wireless access to networks of the future," *IEEE Commun. Mag.*, vol. 25, no. 6, pp. 22-31, June 1987.

[4] *Proc. IEEE* (Special Issue on Hardware and Software for Digital Signal Processing), S. K. Mitra and K. Mondal, Eds., Sept. 1987.

[5] J. Hagenauer, N. Seshadri, and C.-E. W. Sundberg, "The performance of rate-compatible punctured convolutional codes for digital mobile radio," *IEEE Trans. Commun.*, vol. 32, no. 7, pp. 966-980, July 1990.

[6] R. V. Cox, S. L. Gay, Y. Shoham, S. Quackenbush, N. Seshadri, and N. S. Jayant, "New directions in subband coding," *IEEE J. Select. Areas Commun.* (Special Issue on Voice Coding for Communications), vol. 6, no. 2, pp. 391-409, Feb. 1988.

[7] P. Noll, "Effects of channel errors on the signal-to-noise performance of speech-encoding systems," *Bell Syst. Tech. J.*, vol. 54, pp. 1615-1636, Nov. 1975.

[8] N. Rydbeck and C.-E. Sundberg, "Analysis of digital errors in nonlinear PCM systems," *IEEE Trans. Commun.*, vol. COM-24, no. 1, pp. 59-65, Jan. 1976.

[9] R. Steele, C.-E. Sundberg, and W. C. Wong, "Transmission errors in companded PCM over Gaussian and Rayleigh fading channels," *AT&T Bell Lab. Tech. J.*, vol. 63, no. 6, pp. 955-990, July-Aug. 1984.

[10] D. J. Goodman and C.-E. Sundberg, "Transmission errors and forward error correction in embedded differential pulse code modulation," *Bell Syst. Tech. J.*, vol. 62, no. 9, pp. 2735-2766, Nov. 1983.

[11] B. Masnick and J. K. Wolf, "On unequal error protection codes," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 600-607, Oct. 1967.

[12] J. Hagenauer, "Rate compatible punctured convolutional codes and their applications," *IEEE Trans. Commun.*, vol. COM-36, pp. 389-400, Apr. 1988.

[13] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[14] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[15] R. E. Crochiere, R. V. Cox, and J. D. Johnston, "Real-time speech coding," *IEEE Trans. Commun.*, vol. COM-30, pp. 621–634, Apr. 1982.

[16] T. A. Ramstad, "Subband coder with a simple adaptive bit allocation algorithm," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 203–207, Apr. 1982.

[17] M. Honda and F. Itakura, "Bit allocation in time and frequency domain for predictive coding of speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 465–473, June 1985.

[18] F. K. Soong, R. V. Cox, and N. S. Jayant, "A high quality subband speech coder with backward adaptive predictor and optimal time-frequency bit assignment," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 2387–2390, Apr. 1986.

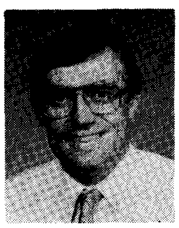
[19] J. Rothweiler, "Polyphase quadrature mirror filters—A new subband coding technique," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 1280–1283, Apr. 1983.

[20] R. V. Cox, "The design of uniformly spaced and nonuniformly spaced pseudoquadrature mirror filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1090–1096, Oct. 1986.

[21] H. Suda and T. Miki, "An error protected 16 kb/s voice transmission for land mobile radio channel," *IEEE J. Select. Areas Commun.*, vol. 6, no. 2, pp. 346–352, Feb. 1988.

[22] J. B. Cain, G. C. Clark, and J. M. Geist, "Punctured convolutional codes for rate $n - 1/n$ and simplfied maximum likelihood decoding," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 97–100, Jan. 1979.

[23] D. J. Goodman and C.-E. Sundberg, "Combined source and channel coding for matching the speech transmission rate to the quality of the channel," *Bell Syst. Tech. J.*, vol. 62, no. 7, pp. 2017–2036, Sept. 1983.



Richard V. Cox (S'69–M'70–SM'87–F'91) received the B.S. degree in 1970 from Rutgers University, Piscataway, NJ, and the M.A. and Ph.D. degrees in 1972 and 1974, respectively, from Princeton University, Princeton, NJ, all in electrical engineering.

From 1973 through 1977 he was a Member of the Technical Staff with the Aerospace Corporation, El Segundo, CA, working in the areas of image processing and queuing theory for operations research. In 1977 he joined the faculty of the Department of Electrical Engineering of Rutgers University, teaching courses and conducting research in the field of digital signal processing. He joined Bell Laboratories in 1979 and has worked in various aspects of speech and audio coding, speech privacy, digital signal processing, and real-time signal processing implementations. Most recently he has been working on robust coding strategies for noisy channels. He is also currently involved with CCITT standardization of speech coding at 16 kb/s. His present position at Bell Labs is Supervisor of the Digital Principles Research Group in the Signal Processing Research Department.

Dr. Cox is Chairman of the Speech Technical Committee and a member of the Administrative Committee for the IEEE Signal Processing Society.

Joachim Hagenauer (M'79–SM'87) received the Ing. (grad.) degree from Ohm-Polytechnic Nurnberg, Germany, in 1963, and the Dipl.Ing and the Dr.Ing. degrees in electrical engineering from the Technical University of Darmstadt, Germany, in 1968 and 1974, respectively.

At Darmstadt University, he served as an Assistant Professor and Docent. From May 1975 to September 1976 he held a postdoctoral fellowship at the IBM T.J. Watson Research Center, Yorktown Heights, NY, working on error-correction coding for magnetic recording. Since 1977 he has been with the German Aerospace Research (DLR), Oberfaffenhofen. From 1980 he was the Head of the Communication Theory Group, and since January 1990 he has been a Director of the Institute of Communications Technology at DLR. During 1986–1987 he spent a sabbatical year as "Otto Lilenthal Fellow" at AT&T Bell Laboratories, Crawford Hill, NJ, working on joint source/channel coding and on trellis coded modulation. He served as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS during 1988–1989. His research interests include convolutional coding, data transmission via fading channels, and mobile communications. He is teaching graduate courses as a part-time professor at the Technical University of Munich.

Nambirajan Seshadri (S'81–M'87) received the B.E. degree in electronics and communication from the University of Madras, India, in 1982, and the M.S. and Ph.D. degrees in computer and systems engineering from the Rensselaer Polytechnic Institute, Troy, NY, in 1984 and 1986, respectively.

He is currently a Member of the Technical Staff at AT&T Bell Laboratories, Murray Hill, NJ, in the Signal Processing Research Department. His research interests include signal processing for speech and data communications.

Carl-Erik W. Sundberg (S'69–M'75–SM'81–F'90) was born in Karlskrona, Sweden, on July 7, 1943. He received the M.S.E.E. and Dr.Techn. degrees from the Lund Institute of Technology, University of Lund, Lund, Sweden, in 1966 and 1975, respectively.

Currently he is a member of the Technical Staff at the Signal Processing Research Department, AT&T Bell Laboratories, Murray Hill, NJ. Before 1976 he held various teaching and research positions at the University of Lund. During 1976, he was with the European Space Research and Technology Center (ESTEC), Noordwijk, The Netherlands, as an ESA Research Fellow. From 1977 to 1984 he was a Research Professor (Docent) in the Department of Telecommunication Theory, University of Lund, Lund, Sweden. He has held positions as Consulting Scientist at LM Ericsson, SAAB-SCANIA, Sweden, and at Bell Laboratories, Holmdel, NJ. His consulting company, SUNCOM, has been involved in studies of error control methods and modulation techniques for the Swedish Defense, a number of private companies, and international organizations. His research interests include source coding, channel coding, digital modulation methods, fault-tolerant systems, digital mobile radio systems, spread-spectrum systems, digital satellite communications systems, and optical communications. He has written over 65 published journal papers and contributed over 90 conference papers. He holds 12 U.S., Swedish, and international patents. He is coauthor of *Digital Phase Modulation* (New York: Plenum, 1986) and *Topics in Coding Theory* (New York: Springer-Verlag, 1989).

Dr. Sundberg was a member of the IEEE European-African-Middle East Committee (EAMEC) of COMSOC from 1977 to 1984. He is a member of COMSOC Communication Theory Committee and Data Communications Committee. He was also a member of the Technical Program Committees for the International Symposium on Information Theory, St. Jovite, Canada, October 1983, and for the International Conference on Communications, ICC'84, Amsterdam, The Netherlands, May 1984. He has organized and chaired sessions at a number of international meetings. He has been a member of the International Advisory Committee for ICCS'88 and ICCS'90 (Singapore). He served as Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 1988–1989. He is a member of Svenska Elektroingenjörers Riksförening (SER) and the Swedish URSI Committee (Svenska Nationalkommitten för Radiovetenskap). In 1986 he and his coauthor received the IEEE Vehicular Technology Society's Paper of the Year Award and in 1989 he and his coauthors were awarded the Marconi Premium Proceedings IEE Best Paper Award.